

สุรจิต ภูมิคง : การออกแบบและพัฒนาขั้นตอนวิธีสำหรับการสกัดข่าวด้านความปลอดภัย
(THE DESIGN AND DEVELOPMENT OF AN ALGORITHM FOR SAFETY-
RELATED NEWS EXTRACTION) อาจารย์ที่ปรึกษา : อาจารย์ ดร.จิตมนต์ อังสกุล,
110 หน้า.

งานวิจัยนี้มุ่งพัฒนาขั้นตอนวิธีการสกัดข่าวด้านความปลอดภัยแบบอัตโนมัติ เพื่อสนับสนุนระบบวางแผนการท่องเที่ยวออนไลน์ภายใต้เงื่อนไขบังคับด้านความปลอดภัย ให้สามารถวิเคราะห์และปรับปรุงข่าวสารด้านความปลอดภัยให้ทันสมัยอยู่เสมอ ขั้นตอนวิธีนี้ทำโดยการดึงข่าวที่เกี่ยวข้องกับความปลอดภัยจากเว็บไซต์รวบรวมข่าว และจัดประเภทของข่าวเป็น 5 ด้าน ได้แก่ ด้านสุขภาพ ด้านการจราจร ด้านภัยธรรมชาติ ด้านความไม่สงบของบ้านเมือง และด้านอาชญากรรม หลังจากนั้นคำศัพท์ที่เกี่ยวข้องกับความปลอดภัยจะถูกสกัดออกจากเนื้อหาข่าว เพื่อนำไปใช้ในการวิเคราะห์ระดับความรุนแรงของเหตุการณ์ และคาดการณ์ระยะเวลาสิ้นสุดของเหตุการณ์ โดยการสกัดคำศัพท์จะใช้ 2 เทคนิค คือขั้นตอนวิธีการสกัดคำศัพท์โดยใช้พจนานุกรมเท่านั้น และขั้นตอนวิธีการสกัดคำศัพท์โดยใช้พจนานุกรมคำศัพท์และหลักการเอ็นแกรม

ผลการทดลองพบว่า ขั้นตอนวิธีการสกัดคำศัพท์โดยใช้พจนานุกรมเท่านั้น มีความถูกต้องในการสกัดข่าวโดยเฉลี่ยทั้ง 5 ด้าน 81.21% และมีค่าความคลาดเคลื่อนในการวิเคราะห์ระดับความรุนแรงของเหตุการณ์โดยเฉลี่ยทั้ง 5 ด้านเท่ากับ 0.37 ในขณะที่ขั้นตอนวิธีการสกัดคำศัพท์โดยใช้พจนานุกรมคำศัพท์และหลักการเอ็นแกรม มีความถูกต้องในการสกัดข่าวโดยเฉลี่ยทั้ง 5 ด้าน 86.01% มีค่าความคลาดเคลื่อนในการวิเคราะห์ระดับความรุนแรงของเหตุการณ์โดยเฉลี่ยทั้ง 5 ด้านเท่ากับ 0.32 ซึ่งแสดงให้เห็นว่า การนำหลักการเอ็นแกรมเข้ามาใช้ ช่วยเพิ่มประสิทธิภาพของขั้นตอนวิธีในการสกัดข่าวและการวิเคราะห์ระดับความรุนแรงของเหตุการณ์ได้

SURAJIT PUMIKONG : THE DESIGN AND DEVELOPMENT OF AN
ALGORITHM FOR SAFETY-RELATED NEWS EXTRACTION. THESIS
ADVISOR : JITIMON ANGSAKUN, D.ENG., 110 PP.

NEWS EXTRACTION/SAFETY-RELATED NEWS/LEVEL OF VIOLENCE

This research aims to develop a safety-related news extraction method for supporting the online travel planner under a safety constraint. The method facilitates the travel planner to analyze and update news automatically. It retrieves safety-related news from news sites and classifies them into five categories: health, traffic, natural disaster, country's unrest and crime. Then the method extracts safety-related keywords from the contents of news in order to apply for analyzing the level of violence of events and predicting the time to finish the events. The keyword extraction applies two techniques that are dictionary based algorithm and *N*-gram based algorithm.

The experimental results illustrate that the dictionary based algorithm gets 81.21% of accuracy in keyword extraction and 0.37 of mean absolute error (*MAE*) in analyzing the level of violence. While the dictionary and *N*-gram based algorithm achieves with 86.01% of accuracy in keyword extraction and 0.32 of *MAE* in analyzing the level of violence. The results indicate that the *N*-gram could improve the performance of algorithm in keyword extraction and analyzing the level of violence.

School of Information Technology

Academic Year 2012

Student's signature _____

Advisor's Signature _____